SET-BASED LABEL PROPAGATION OF FACE IMAGES

Chao Xiong and Tae-Kyun Kim

Department of Electrical and Electronic Engineering, Imperial College London, UK

ABSTRACT

Graph-based Semi-Supervised Learning (SSL) has proven to be an effective tool for label propagation, however, its accuracy is highly dependent on how to form the data weight matrix, in which each element is obtained as the similarity between every pair of data points. Inspired by the success of set-based recognition methods, a novel approach is brought up to incorporate the set-to-set matching as well as single-to-single matching when building up the weight matrix. Canonical Correlation Analysis (CCA), which measures the principal angles between two manifolds, is adopted to compute the set similarity. Moreover, Local Binary Pattern, an effective texture descriptor, is investigated as a data representation to further improve the label propagation performance. The proposed approach is evaluated on two public face image data sets, and shown to significantly outperform the standard SSL methods in terms of accuracy.

Index Terms— Face recognition, Label propagation, Semi-supervised learning, CCA, Local binary pattern

1. INTRODUCTION

Semi-supervised learning (SSL) aims to take labeled data points as seeds and utilize readily available unlabeled data points to improve the learning, or recognition accuracy. Graph-based SSL has been proven efficient in various SSL problems [1, 2, 3, 12]. It creates a graph whose vertices correspond to the labeled and unlabeled data while the edge weights encode the similarity between each pair of data points. For a given weight matrix, estimated labels of those unlabeled is obtained by a closed form solution in the way that data points connected by large weights are given similar data labels. The way to measure data similarities for the weight matrix is crucial to obtain the good label propagation accuracy [1]. Despite a large volume of studies, there has been little attention on how to create the weight matrix. Gaussian Kernel Similarity based measure is most commonly used, and the sparsity induced measurement is recently introduced to improve the label propagation accuracy [1].

The wide utilization of video cameras provides with enormous media resources, which enlightens the importance of set-based recognition methods [4, 6, 7, 8]. A set of images is often formed from a video sequence or long term observations. Rather than taking a single image as input, the methods utilize a set. Capturing and matching set characteristics has shown to improve the recognition performance. Nearest Neighbor [5] is the most straightforward and conducts the comparison in a pairwise manner by directly referring to individual images. More successful approaches try to identify set-characteristics, or capture underlying semantic information of image sets. They include parametric density-based methods and manifold-based methods. The parametric density methods e.g. Kullback-Leibler Divergence (KLD) [4], often fail if training and test data sets do not exhibit strong statistical dependency. Canonical Correlation Analysis (CCA), or called principal angles, which models a set as a manifold and measures the cosine of principal angles between the two manifolds, is widely utilized in recent studies and offers excellent performance [6, 7, 8].

In this paper, we investigate set-matching for graph-based label propagation. Previous studies on Graph-based SSL are mainly conducted by directly invoking a single image as input; thus, do not consider the internal coherent set information. A related study [9] proposes a probabilistic discriminative model for propagating labels by incorporating set-constrains: faces from the same photo must have different identities and faces from a single video track must be of the same person. Our approach shares the similar inspiration, however provides a more general framework of utilizing and matching sets. The standard SSL using the hard-constraints in [9] is implemented and compared in the experiment section 4. This paper presents a novel idea to incorporate set-to-set similarity into single-to-single similarity when depicting the edges on the graph as shown in Figure 1. The integration process emerges in constructing the weight matrix, which is subsequently used for label propagation. The edge weight between sets is computed with the aid of CCA and then combined with single-sample weight in a weighted sum manner so that the constructed graph contains information of two layers.

2. KEY INGREDIENTS OF PROPOSED LEARNING

2.1. Graph Based Semi-Supervised Learning

The label propagation framework in this paper follows [2]. The conceptual graph is model as $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ is

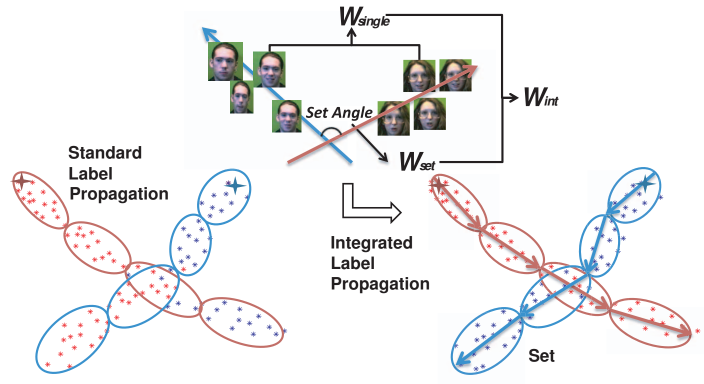

Fig. 1. **Simulated results of the proposed method.** The data has two classes, each class is formed by four sets. Each set contains data points of the same class, shown by a different colored ellipse. There are two labeled data points, annotated as big stars, and the rest unlabeled. The propagated labels are shown by different colored points. Traditional label propagation (left) gets the estimated labels mixed up at the cross section and, from that part, misleads the label propagation, while the proposed method (right) performs the label propagation correctly till the end. In the proposed method, the weight in the graph is computed by combining the single-to-single $\boldsymbol{W}_{single}$ and set-to-set $\boldsymbol{W}_{set}$ similarity so that the label propagation is correctly guided by *set angles*.

a set of nodes (data vectors or images) and $\boldsymbol{E}$ refers to edges that specify proximity between every two nodes. The proximity between two data vectors are measured by

$$\boldsymbol{W}(i,j) = exp\left(-\frac{1}{m}\cdot\sum_{k=1}^{m}\frac{(x_{ik}-x_{jk})^2}{\sigma_k^2}\right) \quad (1)$$

where x_{ik} stands for the k-th component of the i-th vector $\boldsymbol{x}_i \in \mathbf{R}^m$, and σ_k is the length scale hyperparameter along each dimension. Consider that $\boldsymbol{W}$ is segmented in terms of labeled (L) and unlabeled data (U) into four parts, then the estimated labels of unlabeled data points $\boldsymbol{F}_u$ are given in a closed form solution as

$$\boldsymbol{F}_u = (\boldsymbol{D}_{UU} - \boldsymbol{W}_{UU})^{-1}\cdot\boldsymbol{W}_{UL}\cdot\boldsymbol{F}_l \quad (2)$$

where $\boldsymbol{F}_l$ is the label matrix of the labeled data, $\boldsymbol{D}$ is a diagonal matrix s.t. $d_{ii} = \sum_j \boldsymbol{W}(i,j)$.

2.2. Canonical Correlation Analysis

CCA finds optimal coordinate systems for two sets of data vectors so that the corresponding projections are best correlated [7]. With the aid of CCA, the similarity between the two data sets is given as the cosine of the principal angles (also called canonical correlations) of two respective linear subspaces $\boldsymbol{\mathcal{L}}_{t_i}$ and $\boldsymbol{\mathcal{L}}_{t_j}$ as

$$\lambda_{t_i,t_j,k} = cos\theta_k = \max_{\boldsymbol{q}_{t_i,k}\in\boldsymbol{\mathcal{L}}_{t_i}}\ \max_{\boldsymbol{q}_{t_j,k}\in\boldsymbol{\mathcal{L}}_{t_j}} \boldsymbol{q}_{t_i,k}^T\boldsymbol{q}_{t_j,k} \quad (3)$$

where t_i is a set index, $\boldsymbol{Q}_{t_i} = [\boldsymbol{q}_{t_i,1},...,\boldsymbol{q}_{t_i,d}]$ and $\boldsymbol{Q}_{t_j} = [\boldsymbol{q}_{t_j,1},...,\boldsymbol{q}_{t_j,d}]$ are the canonical vectors of $\boldsymbol{S}_{t_i}$ and $\boldsymbol{S}_{t_j}$, subject to $\boldsymbol{q}_{t_i,k}^T\boldsymbol{q}_{t_i,k} = \boldsymbol{q}_{t_j,k}^T\boldsymbol{q}_{t_j,k} = 1$, $\boldsymbol{q}_{t_i,k}^T\boldsymbol{q}_{t_i,l} = \boldsymbol{q}_{t_j,k}^T\boldsymbol{q}_{t_j,l} = 0$, if $k \neq l$ [7].

3. PROPOSED APPROACH

3.1. Problem Formulation

We are given a face image pool composed of both labeled and unlabeled face identities. The labeled images are denoted as ($\boldsymbol{x}_1$,$\boldsymbol{y}_1$), ($\boldsymbol{x}_2$,$\boldsymbol{y}_2$),...,($\boldsymbol{x}_l$,$\boldsymbol{y}_l$), where $\boldsymbol{x}_i$ represents an image vector for the i-th image with the corresponding label $\boldsymbol{y}_i$. Also, there are u unlabeled images $\boldsymbol{x}_{l+1}$,$\boldsymbol{x}_{l+2}$,...,$\boldsymbol{x}_n$, where $n = l + u$. Normally, $l \ll u$. All the image vectors compose the vertex set $\boldsymbol{V}$ in the graph. This image pool also has the notion of image sets, each of which is formed by a small subset of face images with the same identity. Those subsets may be formed by a set of frames in a single video, by a set of long-term observations, or by a random subset. Each image $\boldsymbol{x}_i$ thus belongs to a certain set $\boldsymbol{S}_{t_i}$, where t_i is the corresponding set tag.

3.2. Integrated Label Propagation

Our approach is to take a set information into account for propagating labels from the known to unknown. Most previous studies utilize merely single-to-single matching, and fail to propagate labels when face images of the same identity exhibit significant appearance changes. Instead, set-to-set matching captures underlying data manifolds and achieves more robust recognition performance by comparing a manifold to a manifold. Note that varying face images of the same person are often constrained on a low-dimensional manifold or subspace. However, those set-based methods lose accuracy if the data variation in each set is limited. We propose the integrated label propagation which combines the set-to-set similarity with the single-to-single measurement at the stage of forming the weight matrix. This way, the edges $\boldsymbol{E}$ in the newly formed graph reveal the two layers of information. The weight matrix in the proposed method is given as

$$\boldsymbol{W}_{int} = (1-w)\cdot\boldsymbol{W}_{single} + w\cdot\boldsymbol{W}_{set} \quad (4)$$

where $\boldsymbol{W}_{single}$ is inherited from the traditional SSL (1) and $\boldsymbol{W}_{set}$ is the corresponding set-based weight matrix, and w is the weight parameter determining the relative importance of $\boldsymbol{W}_{single}$ to $\boldsymbol{W}_{set}$.

Canonical Correlation Analysis [8] applies the maximal correlations for any pair of sets. We consider that each correlation coefficient is of a certain importance and is retained.

Correlation coefficients reflect the relation between two data sets along the common modes of data variations, and thus reveal the degree of similarity between the two sets. Our empirical observations are also that the face image sets with the same identity tend to return much larger coefficients on average compared to those of different identities, especially for the first several coefficients. A formula to derive the set-to-set weight matrix is given as

$$W_{set}(i,j) = exp\left(-\frac{1}{d}\cdot\sum_{k=1}^{d}\frac{(\lambda_{t_i,t_j,k}-1)^2}{\sigma^2}\right) \tag{5}$$

where $W_{set}(i,j)$ represents the edge weight between two samples with the set tags as t_i and t_j (i.e. $\boldsymbol{x}_i \subseteq \boldsymbol{S}_{t_i}$, $\boldsymbol{x}_j \subseteq \boldsymbol{S}_{t_j}$). $\lambda_{t_i,t_j,k}$ stands for the k-th correlation coefficient derived along the corresponding dimension by (3). δ, here, is also a length scale hyperparameter as defined in the traditional way of constructing a weight matrix. Since the set weight is summed with the original weight matrix $\boldsymbol{W}_{single}$, the numerical distribution of $\boldsymbol{W}_{set}$ should be similar to that of $\boldsymbol{W}_{single}$. Note that the Gaussian based measurement above is similar to (1). The smoothing function is monotonically increasing with increasing gradients in terms of the correlation coefficients within [0,1]. The boundary value is 1 for $\lambda = 1$ and a minimal value for $\lambda = 0$.

3.3. Further Improvement by Multi-Scale and Multi-Block Local Binary Pattern

A choice of face image representation i.e. the vertex $\boldsymbol{V}$ in the graph, also determines label propagation accuracy. Raw-pixel image representation is quite vulnerable even to monotonic intensity value changes. Local binary pattern, a powerful texture descriptor, has gained an increasing attention for face recognition. It has demonstrated reasonable robustness to image appearance changes due to illumination and pose variations, while it largely reduces the data size [10]. We adopt an extension of LBP - Multi-scale (MSLBP) and Multi-block LBP (MBLBP) [11] to encode the facial geometrical information. The image representation is given as

$$\boldsymbol{x}_i = \begin{matrix}[\boldsymbol{H}^{B_1}_{P,R_1}, & \boldsymbol{H}^{B_1}_{P,R_2}, & \dots & \boldsymbol{H}^{B_1}_{P,R_N}, & \\ \dots & \boldsymbol{H}^{B_M}_{P,R_1}, & \boldsymbol{H}^{B_M}_{P,R_2}, & \dots & \boldsymbol{H}^{B_M}_{P,R_N}]^T\end{matrix} \tag{6}$$

where $\boldsymbol{H}^{B_j}_{P,R_i}$ represents the LBP histogram (LBPH) of the sub-block B_j with P neighboring points laying on a circle of radius R_i. See [11] for the details.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Database. The proposed method has been evaluated on the two public face data sets: Univ. of Essex (Faces95) and FEI Face database. In Face95 database, each individual has 20 face images derived from video clips of multiple races. Images exhibit appearance changes due to beard, glasses, accessories, or head poses. It contains 1440 face images of 200×180 pixel resolutions. For FEI database, images were taken against a white homogeneous background in an upright frontal position with profile rotations of up to 90 degrees. There are 14 images for each of 200 individuals, a total of 2800 images in 640×480 pixels. All images were resized into 46×56 pixels and processed by Histogram Equalisation and Gaussian smoothing for raw pixel presentations.

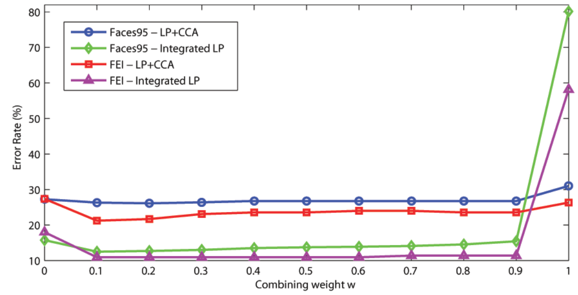

Fig. 2. Effect of the weight parameter on label propagation accuracy. LP+CCA is the proposed method with raw pixels and Integrated LP the method with LBP.

Settings. The databases considered do not include set tags. We therefore formed sets by randomly drawing small subsets of face images within the same identity. In Faces95, there were formed 6 sets of unlabeled images per person, while 4 sets in FEI database. When more videos or long-term observations are available, the sets may be formed using such a group information. For a fair comparison with the traditional LP method, the single-to-single weight matrix $\boldsymbol{W}_{single}$ in the traditional LP method is adjusted such that the given set information is exploited (called LP+Set). The weight between instances within the same set is set as 1, otherwise by (1). This implementation is similar to [9] where the hard set-constrains were used.

The length scale hyperparameters were set following [12, 3]. 1/3 of the minimal between-class distance was used so that the weight of this edge is nearly 0. For LBP, the number of neighbors was set to 8, and three scales were included with the radius as 1, 2 and 3, respectively. Each image is firstly segmented into several sub-regions (5×3 for Face95, 16×12 for FEI) before applying MSLBP. The uniform pattern is adopted in LBPH to include more stable and robust patterns under geometric transformation as in [11].

Results on the effect of combining weight w. The parameter w determines the relative importance between the two weight matrices. Figure 2 shows the effect of changing the weight on label propagation accuracy. All the results were obtained using one labeled image per subject. The optimal weight assignment demonstrates a lower error rate than solely relying on either the single-to-single $(w = 0)$ or set-to-set

Faces95

Trial	LP	LP+SET	LP+CCA	LP+LBP	I-LP
1	50.5	27.27	26.1	15.76	13.01
2	49.42	10.23	8.33	9.21	8.16
3	45.47	18.57	15.94	9.7	8.23

FEI

Trial	LP	LP+SET	LP+CCA	LP+LBP	I-LP
1	46.31	27.38	21.23	18	10.92
2	38	28.15	23.38	12.92	3.85
3	35.23	26	25.69	6.92	3.85

Table 1. Comparison among different approaches in terms of label propagation error rate: LP shows the result of the traditional LP with raw pixels; LP+Set is the traditional LP with the weight of the same set regulated as 1; LP+CCA is the proposed method with raw pixels; LP+LBP follows the setting of LP+Set with LBP; I-LP stands for the integrated label propagation that combines LP, CCA and LBP all together.

$(w = 1)$ based, as the weighted combination counteracts uncommon similarities returned by any of the two methods. The weighted combination reduced the average error rate by 3%.

Comparison between the integrated and traditional label propagation. The previous experiments demonstrate that the optimal selection of the weight parameter may vary in different situations - there is no uniform answer to the setting. In the following experiments, the combining weights are inherited from the optimal value obtained in Figure 2: $w = 0.2$ for LP+CCA, $w = 0.1$ for Integrated LP (Faces95); while $w = 0.1$ for LP+CCA, $w = 0.5$ for Integrated LP (FEI). Since the formation of sets was done in a randomized way, label propagation accuracy might vary in each run. In the experiment, three different trials with same settings were conducted for Faces95 or FEI dataset.

As observed in Table 1, the performance of the standard Label Propagation with direct reference to set information, i.e. LP+Set, led to significant improvement on accuracy comparing to traditional LP. The label propagation accuracy was improved by the proposed method in each experiment for both raw pixel or LBP representations. The improvement is about 5% on average over the standard LP+Set with raw-pixels where the weight computation is quite unreliable due to variations of illumination, pose, and etc. For faces95, which includes illumination changes and geometrical displacements, using Local Binary Pattern solely brought up about 12% error rate reduction. Over the LP method with LBP, the proposed integration method further led to the meaningful improvement in terms of accuracy. Overall, in the methods using single data points, if certain nodes are mistakenly labeled, other adjacent nodes also share the ”false” labels. In contrast, with the proposed set-based method, the manifold structure of image sets revealed by CCA counteracted the ”uncommon” closeness in Euclidean space in LP+SET leading to the improvement in accuracy.

5. CONCLUSION

The proposed method has incorporated set-information and Local Binary Pattern into the conventional graph-based framework for label propagation. The basic idea is to combine the set-to-set and single-to-single similarity measurements, which takes place when forming the edge weight matrix for the graph. The set-to-set distance is measured by evaluating the correlation coefficients obtained by CCA with a Gaussian smoothing function and then weighted summed to the original single-to-single weight matrix. The combination has shown the significant improvement in the label propagation accuracy on the two public benchmarks. Also, the multi-scale and multi-block Local Binary Pattern replacing the raw pixel representation demonstrated the huge accuracy gain. Future research will focus on better alternatives of integration, set similarity measurement, feature extraction as well as an automatic way of setting the combining weight.

6. REFERENCES

[1] H.Cheng, Z.Liu and J.Yang, “Sparity induced similarity measure for label propagation,” *ICCV*, 2009.

[2] X. Zhu, Z. Ghahramani and J. Lafferty, “Semi-supervised learning using gaussian fields and harmonic functions,” *ICML*, 2003.

[3] X.H. Zhang and W.S.Lee, “Hyperparameter learning for graph based semi-supervised learning algorithms,” *NIPS*, 2007.

[4] G. Shakhnarovich, J.W. Fisher and T. Darrel, “Face recognition from long-term observations,” *ECCV*, 2002.

[5] S. Satoh, “Comparative evaluation of face sequence matching for content-based video access,” *AFG*, 2000.

[6] L. Wolf and A. Shashua, “Learning over sets using kernel principal angles,” *J. Machine Learning Research* 4(10):913–931, 2003.

[7] T.K.Kim, J.Kittler and R.Cipolla, “Discriminative learning and recognition of image set classes using canonical correlations,” *TPAMI* 29(6), 2007.

[8] O. Yamaguchi, K. Fukui, and K. Maeda, “Face recognition using temporal image sequence,” *AFG*, 1998.

[9] A. Kapoor, G. Hua, A. Akbarzadeh and S. Baker, “Which faces to tag: Adding prior constraints into active learning,” *ICCV*, 2009.

[10] T. Ojala, M. Pietikinen and T. Menp, “Multiresolution grayscale and rotation invariant texture classification with local binary patterns,” *PAMI*, 24(7):971–987, 2002.

[11] C.H. Chan, “Multi-scale local binary pattern histogram for face recognition,” *PhD Thesis, University of Surrey*, 2008.

[12] X. Zhu and Z. Ghahramani, “Learning from labeled and unlabeled data with label propagation,” *CMU-CALD-02-107*, 2009.